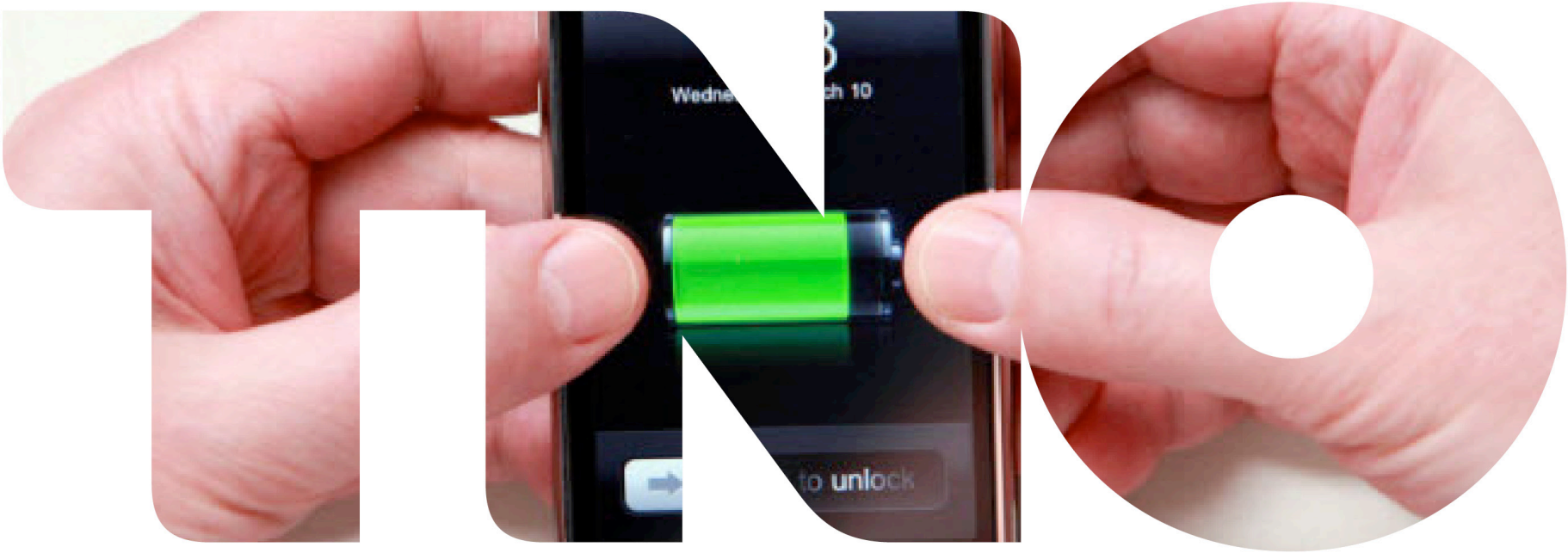
TNO innovation for life

Linking Open Energy Data

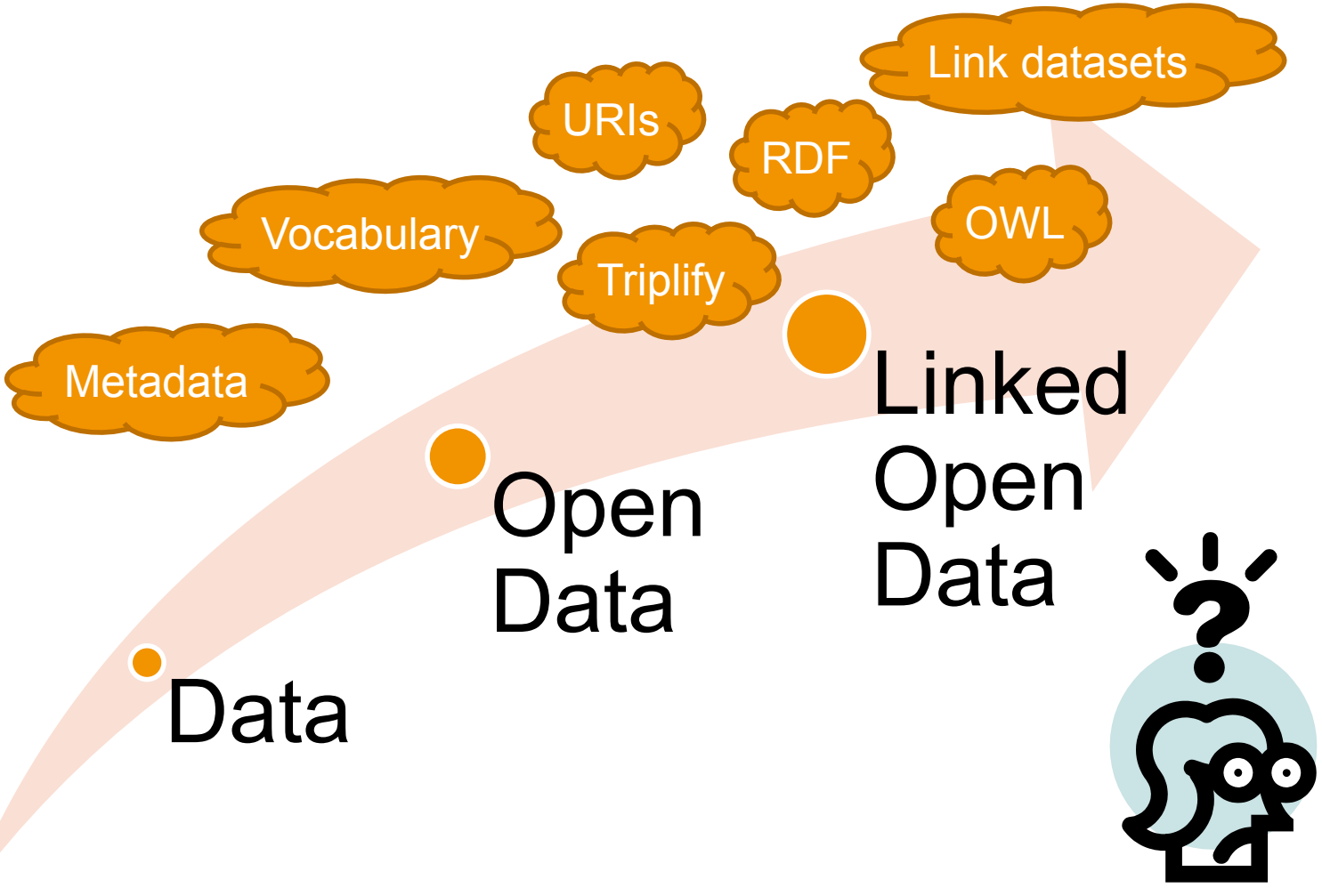
A step-by-step Guide to Publishing Linked Data

Maarten Steen



TNO innovation for life

So you have some data to share…

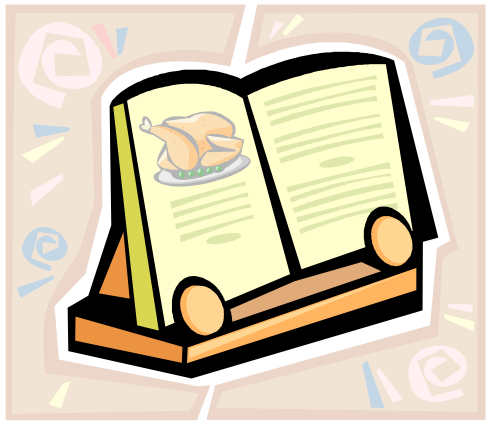


For more concepts and best practices see: http://www.pilod.nl/wiki/Links

TNO innovation for life

A step-by-step Guide to Publishing Linked Data

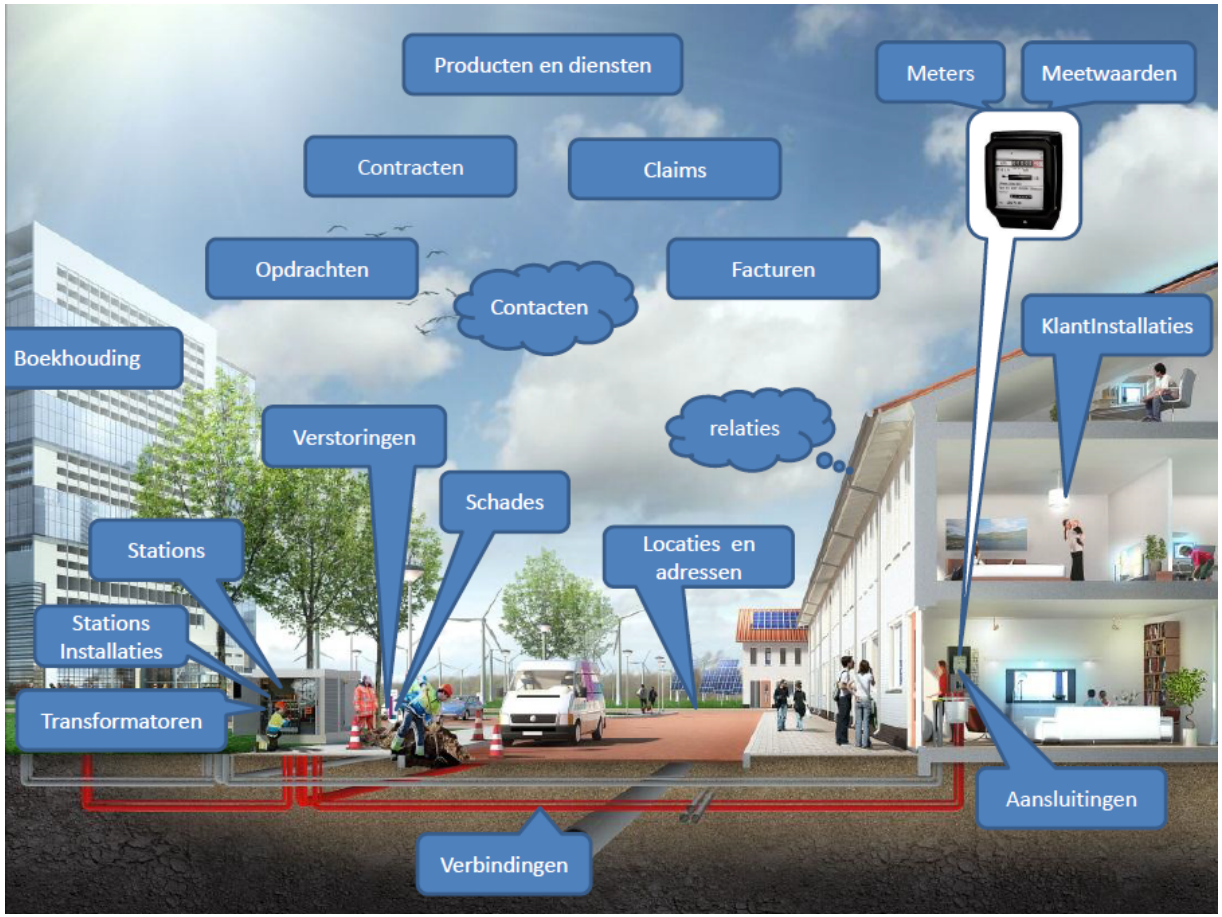
- Step 1: Select data
- Step 2: Prepare the data
- Step 3: Model the data
- Step 4: Define a naming scheme
- Step 5: Convert the data
- Step 6: Organize governance
- Step 7: Add metadata
- Step 8: Publish the data
- Step 9: Link the data



http://www.pilod.nl/wiki/BoekTNO/stappenplan

TNO innovation for life

An Example: the Liander Open Dataset



Step 2:
Prepare data

TNO innovation for life

Groomed data
(Filtered, Anonymized, Aggregated)

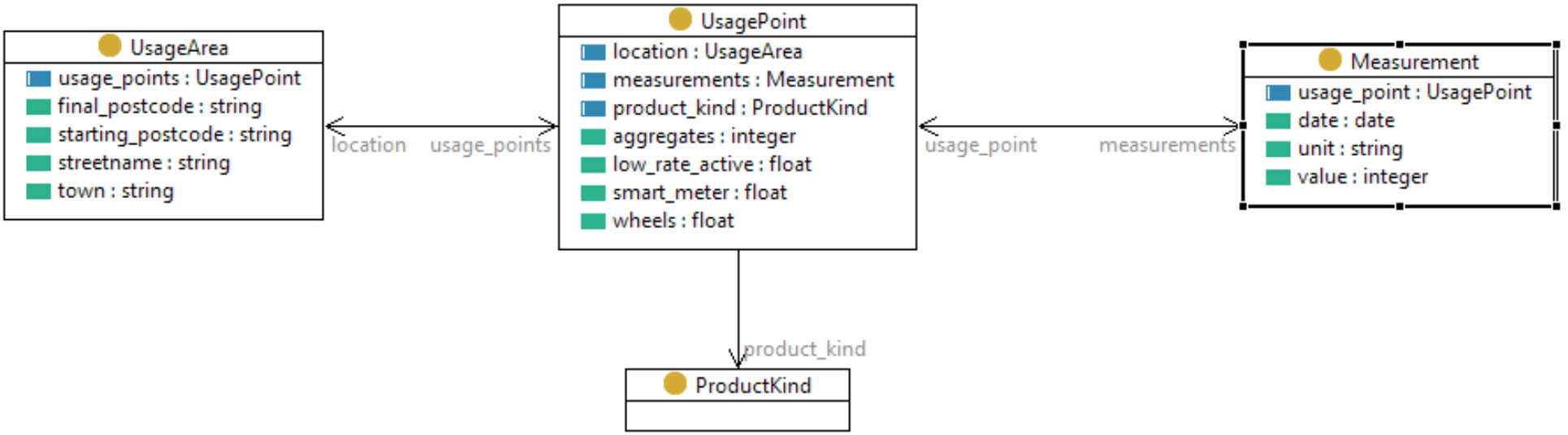
STRAATNAAM	PC_VAN	PC_TOT	WOONPLAATS	PROD	#	SJV	%laag tarief	%Slimme Meter	Avg #Meters
Rijksweg A44	1000AA	1011AB	NIEUW VENNEP	ELK	31	16245	38,71	16,13	1,3
De Ruyterkade	1011AC	1011AC	AMSTERDAM	ELK	32	11433	28,13	15,63	1,1
De Ruyterkade	1011AC	1011AC	AMSTERDAM	GAS	22	3938	0	18,18	
Oosterdokskade	1011AD	1011AD	AMSTERDAM	ELK	20	3665	45	0	1,5
Oosterdokskade	1011AD	1011AE	AMSTERDAM	GAS	11	1248	0	9,09	
Oosterdokskade AB	1011AE	1011AE	AMSTERDAM	ELK	18	3117	33,33	5,56	1
Prins Hendrikkade	1011AG	1011AG	AMSTERDAM	ELK	12	20056	58,33	25	1,6
Prins Hendrikkade	1011AG	1011AH	AMSTERDAM	GAS	27	2199	0	14,81	
Prins Hendrikkade	1011AH	1011AH	AMSTERDAM	ELK	21	3455	14,29	0	1,1
Prins Hendrikkade	1011AJ	1011AJ	AMSTERDAM	ELK	19	8680	26,32	26,32	1
Prins Hendrikkade	1011AJ	1011AJ	AMSTERDAM	GAS	15	1722	0	33,33	

Tools: Database, Excel

Step 3:
Model data

TNO innovation for life

A simple ontology for energy data



Tools: TopBraid Composer

TNO innovation for life

Define a naming scheme (URI strategy)

- For concepts in the ontology:
- **http://{domain}/def/{ontology}#{concept}**

- For objects in the dataset:
- **http://{domain}/{dataset}/id/{concept}/{reference}**

- Examples:
- `<http://data.liander.nl/def/liander>` for the ontology.
- `<http://data.liander.nl/def/liander#UsageArea>` for the concept of Usage Area within the ontology.
- `<http://data.liander.nl/liander/id/UsageArea/7231JS7231JT>` for the Usage Area that starts at postcode 7231JS and ends at postcode 7231JT.

TNO innovation for life

Convert to RDF

- Method 1: Program a script to export data from DB to Turtle syntax
- Method 2: Import in LODRefine and write RDF templates in Google Refine templating language
- Method 3: Define and execute mappings between ontology and database in R2RML or similar

TNO innovation for life

Convert to RDF

- Method 1: Program a script to export data from DB to Turtle syntax

```
psql -A -t -d cerise -o emeter5.nt -c "select '<http://
lod.geodan.nl/cerise-sg/ebif/julianadorp/emeter_' || k.naam || '>
<http://www.w3.org/ns/locn#measurement> <http://lod.geodan.nl/
cerise-sg/ebif/julianadorp/meting_elec_' || k.naam || '_' || e.id ||
'> .' from electralevering e join klant k on e.klant = k.id where
e.waarde is not null;"
```

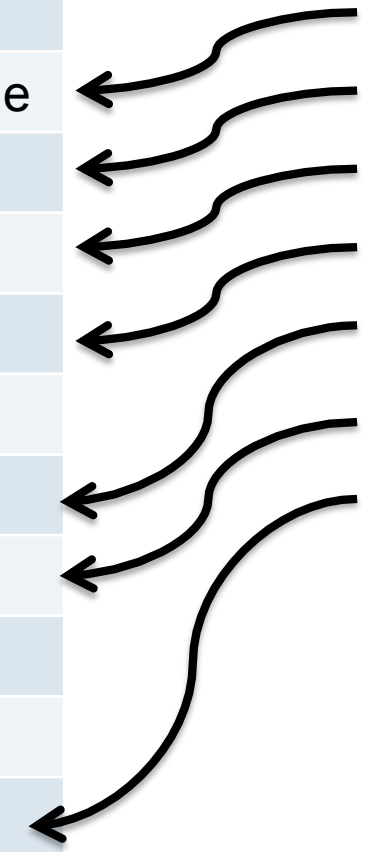
- Method 2: Import in LODRefine and write RDF templates in Google Refine templating language
- Method 3: Define and execute mappings between ontology and database in R2RML or similar

Step 5:
Convert to RDF

TNO innovation for life

Define Mappings with Database Schema

Class	Property
UsageArea	URI
	starting_postcode
	final_postcode
	streetname
	town
UsagePoint	URI
	product_kind
	aggregates
	…
Measurement	URI
	value



Column
POSTCODE_VAN
POSTCODE_TOT
STRAATNAAM
WOONPLAATS
PRODUCTSOORT
AANTAL_AANSLUITING
SJV
…
…

Step 5:
Convert to RDF

TNO innovation for life

Formalize Mappings in R2RML or similar

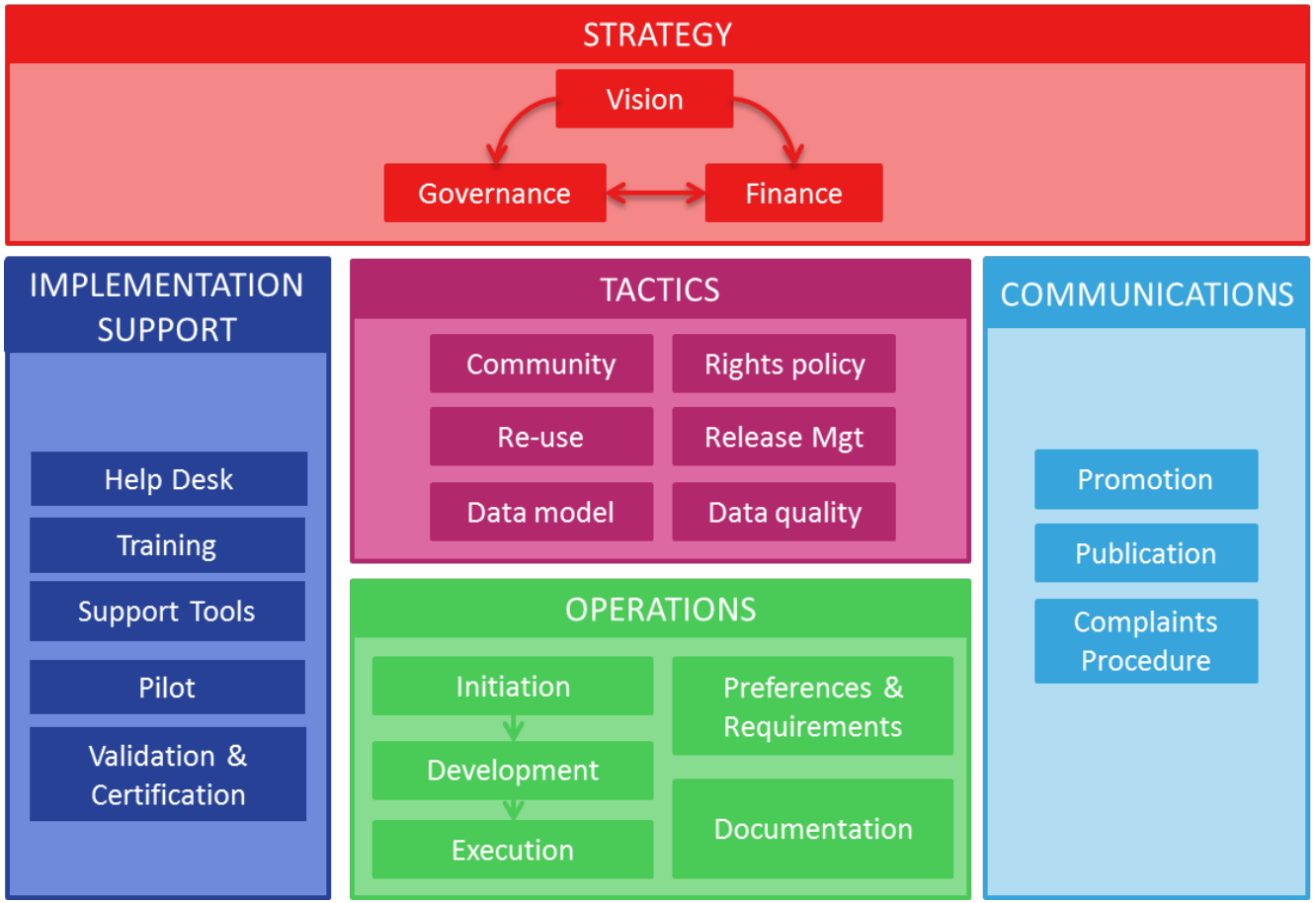
SQL	Target triples
`SELECT DISTINCT postcode_van, postcode_tot FROM kleinverbruik`	`data:UsageArea/{postcode_van}{postcode_tot} a :UsageArea .`
`SELECT DISTINCT postcode_van, postcode_tot, productsoort FROM kleinverbruik WHERE productsoort = 'GAS'`	`data:UsagePoint/{postcode_van}{postcode_tot}G a :UsagePoint ; :location data:UsageArea/{postcode_van}{postcode_tot} ; :product_kind :Gas .`
`SELECT DISTINCT postcode_van, postcode_tot, straatnaam, woonplaats FROM kleinverbruik`	`data:UsageArea/{postcode_van}{postcode_tot} :starting_postcode {postcode_van} ; :final_postcode {postcode_tot} ; :streetname {straatnaam} ; :town {woonplaats} .`
where	`@prefix : <http://data.liander.nl/def/liander#> .` `@prefix data: <http://data.liander.nl/liander/id/> .`

Tools: ontopPro plugin for Protégé

TNO innovation for life

Organize governance using BOMOD

(Management and Development Model for Open Data)



TNO innovation for life

Add metadata

- In this step we make the dataset self-describing by adding metadata. We can do this by simply adding triples to the RDF version of the dataset obtained in step 5 that describe facts about the dataset itself.

```
<http://data.liander.nl/Liander> rdf:type void:DataSet
<http://data.liander.nl/Liander> dct:modified “2014-05-27”
<http://data.liander.nl/Liander> dct:title
        "Liander energy usage dataset."
<http://data.liander.nl/Liander> dct:description
        “Standardized annual energy usage of small users in the
        Liander domain aggregated per postcode area."
<http://data.liander.nl/Liander> dct:creator
        <http://nl.dbpedia.org/resource/Alliander>
<http://data.liander.nl/Liander> dct:date “2014-03-08”
<http://data.liander.nl/Liander> dct:publisher
        <http://www.liander.nl/>
<http://data.liander.nl/Liander> void:vocabulary
        <http://data.liander.nl/def/energy>
```

TNO innovation for life

Publish the data

- Option A:
  - Publish the dataset as a flat file (in RDF/XML or Turtle)
- Option B:
  - Store the data in a triple store and serve it through a SPARQL-endpoint
- Option C:
  - Use R2RML mappings to create a virtual triple store on a relational database
- Option D:
  - Create a REST-style API on top of a triple store

Tools: ontop Quest SPARQL-endpoint through Sesame

TNO innovation for life

Link the data

- Links to ontologies (see step 3)
- Links to other resources
  - Add triples to the dataset
  - Create a separate link set
- For example,
  - Link Usage Areas to postcode areas
  - Add Energy Labels to Usage Areas

TNO innovation for life

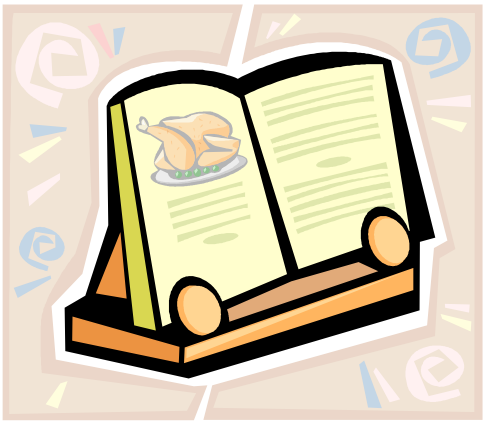
Technology used

Technology	Purpose
Postgres	Relational Database
JDBC	Connect to legacy databases
TopBraid Composer	Create, edit and relate ontologies
Protégé with ontopPro plugin	Create and test DB to ontology mappings
Sesame with ontop Quest	Convert data to RDF; provide SPARQL-endpoint

TNO innovation for life

Summary

- Step 1: Select data
- Step 2: Prepare the data
- Step 3: Model the data
- Step 4: Define a naming scheme
- Step 5: Convert the data
- Step 6: Organize governance
- Step 7: Add metadata
- Step 8: Publish the data
- Step 9: Link the data



http://www.pilod.nl/wiki/BoekTNO/stappenplan

TNO innovation for life

Credits

- Co-authors: Silja Eckartz, Erwin Folmer (TNO)
- Cerise SG project: www.cerise-project.nl
- Alliander: www.liander/nl/opendata
- PLDN Community and Wiki